

Value-Added Source Decomposition

The matrix algebra behind `va_decomposition_workflow.py`

Pakistan vs. Developing Asia, ADB Input–Output Tables (2022)

Companion note for anyone running `va_decomposition_workflow.py`

This note derives the decomposition that the workflow computes: for each sector we attribute 100% of its *gross output value* to the domestic value added created in every source sector, plus imported-input content, plus net product taxes. The method is the standard demand-side Leontief model used throughout the global value-chain (GVC) trade literature; see the references at the end.

1 Where this comes from: the Leontief model

An input–output table can be read in two directions, and the whole method rests on the difference between them.

Reading *across a row* i is the market-clearing (quantity) balance: sector i 's output is split between sales as intermediate inputs to other sectors and sales to final demand f_i ,

$$x_i = \sum_j Z_{ij} + f_i, \quad \implies \quad \mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{f}, \quad \mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{f}, \quad (1)$$

using $Z_{ij} = A_{ij}x_j$ (defined below). Reading *down a column* j is the cost/value balance: the value of sector j 's output equals its intermediate-input bill plus payments to primary factors (value added), imported inputs, and net taxes – this is the accounting identity (2) in the next section.

The behavioural content is the *Leontief production function*: each sector produces with inputs in fixed proportions, so making one unit of output j requires a fixed amount $A_{ij} = Z_{ij}/x_j$ of every input i , independent of scale and with no substitution between inputs. That single assumption turns the A_{ij} into structural technology parameters and makes both balances linear. From the row balance, inverting $\mathbf{I} - \mathbf{A}$ gives the Leontief inverse $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$: it converts final demand into the gross output needed to support it, counting the direct requirement and every upstream round. We do not use the row balance directly – our object lives on the *column* (value) side – but this is where $\mathbf{L}$ earns its meaning as the total-requirements matrix.

The decomposition then asks a value-side question: of the money paid for one unit of sector j 's output, how much ultimately accrues as value added created in each source sector i , versus leaking to imported inputs or taxes? Since value added is the primary-factor share $v_j = va_j/x_j$ left in each sector after its intermediate bill, propagating v_j through $\mathbf{L}$ answers exactly that. Sections 2–4 make this precise; Section 5 compares Pakistan to a developing-Asia benchmark.

2 Data and notation

For the n producing sectors of one country, the ADB table (sheet **Table 1.17**) provides:

Symbol	Dimension	Meaning
$\mathbf{Z}$	$n \times n$	intermediate transactions: Z_{ij} = output of i used as input by j
$\mathbf{x}$	$n \times 1$	gross output; $\hat{\mathbf{x}} = \text{diag}(\mathbf{x})$
$\mathbf{va}^\top$	$1 \times n$	value added at basic prices
$\mathbf{im}^\top$	$1 \times n$	imported-input content
$\mathbf{tax}^\top$	$1 \times n$	net taxes (taxes less subsidies on products)

Let $\mathbf{1}$ denote the column vector of ones. Each sector's output value is exhausted by intermediate purchases, value added, imports and net taxes, giving the column accounting identity

$$x_j = \sum_{i=1}^n Z_{ij} + va_j + im_j + tax_j, \quad \text{i.e.} \quad \mathbf{x}^\top = \mathbf{1}^\top \mathbf{Z} + \mathbf{va}^\top + \mathbf{im}^\top + \mathbf{tax}^\top. \quad (2)$$

3 Technical coefficients and the Leontief inverse

The technical-coefficient matrix normalises each intermediate flow by the *total output* of the *using* sector,

$$\mathbf{A} = \mathbf{Z} \hat{\mathbf{x}}^{-1}, \quad A_{ij} = \frac{Z_{ij}}{x_j}, \quad (3)$$

so A_{ij} is the amount of input i required per unit of sector j 's output – the fixed-proportions Leontief production function. This output denominator (not an “intermediates” denominator) is what makes the total-requirements matrix, the *Leontief inverse*,

$$\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots, \quad (4)$$

count all direct and indirect rounds of production.¹ Define the per-unit-of-output coefficient vectors

$$\mathbf{v}^\top = \mathbf{va}^\top \hat{\mathbf{x}}^{-1}, \quad \mathbf{m}^\top = \mathbf{im}^\top \hat{\mathbf{x}}^{-1}, \quad \mathbf{t}^\top = \mathbf{tax}^\top \hat{\mathbf{x}}^{-1}. \quad (5)$$

Closure. Post-multiplying (2) by $\hat{\mathbf{x}}^{-1}$ and using (3)–(5),

$$\mathbf{1}^\top \mathbf{A} + \mathbf{v}^\top + \mathbf{m}^\top + \mathbf{t}^\top = \mathbf{1}^\top, \quad \implies \quad \mathbf{v}^\top + \mathbf{m}^\top + \mathbf{t}^\top = \mathbf{1}^\top (\mathbf{I} - \mathbf{A}). \quad (6)$$

The column sums of $\mathbf{A}$, plus the value-added, import and tax coefficients, add to one. This identity holds *only* with the output denominator of (3); it is the engine that makes every column of the decomposition sum to 100%.

4 Source decomposition of gross output value

One unit of sector j 's output embodies, through all upstream rounds, domestic value added created in each source sector i :

$$\mathbf{E} = \hat{\mathbf{v}} \mathbf{L}, \quad E_{ij} = v_i L_{ij} \quad (\hat{\mathbf{v}} = \text{diag}(\mathbf{v})), \quad (7)$$

together with embodied imported inputs² and embodied net taxes,

$$\boldsymbol{\mu}^\top = \mathbf{m}^\top \mathbf{L}, \quad \boldsymbol{\tau}^\top = \mathbf{t}^\top \mathbf{L}. \quad (8)$$

Stack these into a single $(n+2) \times n$ matrix whose rows are sources and whose columns are using sectors:

$$\mathbf{G} = \begin{bmatrix} \mathbf{E} \\ \boldsymbol{\mu}^\top \\ \boldsymbol{\tau}^\top \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{v}} \mathbf{L} \\ \mathbf{m}^\top \mathbf{L} \\ \mathbf{t}^\top \mathbf{L} \end{bmatrix}. \quad (9)$$

¹The Neumann series $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots$ converges iff the spectral radius $\rho(\mathbf{A}) < 1$. Here that is guaranteed by the Hawkins–Simon / productivity condition: every sector has strictly positive value added, so by (6) each column of $\mathbf{A}$ sums to $1 - v_j - m_j - t_j < 1$, and $\rho(\mathbf{A}) \leq \|\mathbf{A}\|_1 = \max_j (\text{column sum}) < 1$.

²With a single national table, imports enter as a primary-input-like row and are pushed through the *domestic* inverse $\mathbf{L}$ – the standard non-competitive-imports treatment. “Imported inputs” is therefore a *terminal* source category: we measure how much of each sector's output value leaks abroad, but not the foreign sector in which that value originates. Tracing foreign content further would require a multi-region table (MRIO), as in Koopman, Wang & Wei (2014).

Each column sums to one. Summing the domestic block, $\mathbf{1}^\top \mathbf{E} = \mathbf{1}^\top \hat{\mathbf{v}} \mathbf{L} = \mathbf{v}^\top \mathbf{L}$, so the column totals of $\mathbf{G}$ are

$$\mathbf{v}^\top \mathbf{L} + \mathbf{m}^\top \mathbf{L} + \mathbf{t}^\top \mathbf{L} = (\mathbf{v}^\top + \mathbf{m}^\top + \mathbf{t}^\top) \mathbf{L} \stackrel{(6)}{=} \mathbf{1}^\top (\mathbf{I} - \mathbf{A}) (\mathbf{I} - \mathbf{A})^{-1} = \mathbf{1}^\top. \quad (10)$$

Every sector's output value is fully and exactly attributed to domestic value added by source, imported inputs, and net taxes.

5 Aggregation: 35 ADB sectors $\rightarrow$ 15 paper sectors

Sources and using sectors aggregate differently. Let $g(\cdot)$ map each ADB sector to one of the $G = 15$ paper sectors.

Source rows – summed. With the $\{0, 1\}$ selector $\mathbf{S} \in \{0, 1\}^{G \times n}$, $S_{g,j} = 1 \iff g(j) = g$, the domestic block's rows are summed within each group: $\mathbf{S} \mathbf{E}$. (Imports and net taxes are single rows and carry over unchanged.)

Using columns – output-weighted average. Let $\mathbf{W} \in \mathbb{R}^{n \times G}$ with

$$W_{j,h} = \begin{cases} \frac{x_j}{\sum_{k: g(k)=h} x_k}, & g(j) = h, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

so each column of $\mathbf{W}$ holds output shares that sum to one. The aggregated $(G+2) \times G$ matrix is

$$\mathbf{F} = \begin{bmatrix} \mathbf{S} \mathbf{E} \\ \boldsymbol{\mu}^\top \\ \boldsymbol{\tau}^\top \end{bmatrix} \mathbf{W}. \quad (12)$$

Row-summing (via $\mathbf{S}$) only regroups entries within a column, and each aggregated column is a convex combination (weights sum to one) of columns that already sum to one by (10); hence *every column of $\mathbf{F}$ still sums to exactly one*. The final renormalisation to percentages,

$$\mathbf{M} = 100 \cdot \mathbf{F} \text{diag}(\mathbf{1}^\top \mathbf{F})^{-1}, \quad (13)$$

is therefore an identity up to floating-point error, not a correction. Two zero-output cases must be separated. Individual sectors with $x_j = 0$ are dropped *before* $\mathbf{A}$ is formed (the code keeps only $x > 0$), so $\mathbf{A}$ and $\mathbf{L}$ are never contaminated by a division by zero. The remaining case is a whole paper *group* all of whose member sectors have zero output: then that column of $\mathbf{F}$ is empty and (13) returns an undefined (NaN) column. In this dataset that happens only for the Maldives (Petroleum & Chemicals and Machinery & Electrical); the Maldives is excluded from the benchmark and is not the focal country, and the cross-country median in (14) uses a NaN-robust operator, so such a column never reaches a figure. The entry M_{ij} is the percentage of using-sector j 's gross output value that originates in source i (a paper sector, imported inputs, or net taxes), with columns summing to 100.

Why aggregate the results, not the table. Everything above is computed at the full 35-sector resolution, and only the resulting matrix is aggregated at the very end. Collapsing $\mathbf{Z}$ to 15 sectors first and recomputing $\mathbf{A}$ there would force a single technical coefficient onto each merged group and generally yield different, biased multipliers – the classic input-output aggregation problem. Decomposing at full resolution and aggregating the *results* avoids it.

6 Cross-country benchmark and the difference figure

Applying (3)–(13) to every country c yields $\mathbf{M}^{(c)}$. The developing-Asia benchmark is the *cell-wise median* over the benchmark set $\mathcal{C}$ (developed economies and structural outliers excluded; Pakistan excluded from its own benchmark):

$$M_{ij}^{\text{bench}} = \text{median}_{c \in \mathcal{C}} M_{ij}^{(c)}. \quad (14)$$

The median is *not* additive, so the columns of $\mathbf{M}^{\text{bench}}$ need not sum to 100. The comparison figure plots the difference

$$\Delta = \mathbf{M}^{\text{Pak}} - \mathbf{M}^{\text{bench}}, \quad (15)$$

where $\Delta_{ij} > 0$ (red) means source i contributes a larger share of sector j ’s output value in Pakistan than in the typical developing-Asia economy, and $\Delta_{ij} < 0$ (blue) the reverse.

Because the benchmark columns need not total 100, the columns of Δ need not total zero. A column can therefore read as net red simply because the benchmark medians in it sum to less than 100, not because Pakistan’s sectoral structure is unusual; each cell should be read on its own rather than balancing red against blue within a column. Renormalising every benchmark column to 100 before differencing would restore $\sum_i \Delta_{ij} = 0$, but at the cost of distorting the individual median cells, so we keep the raw medians undistorted.

Equation $\leftrightarrow$ code map

Equation	Code (<code>va_decomposition_workflow.py</code>)	Object
(2)	<code>read_iot(Z, VA, X, IM, TAX)</code>	data
(3)	<code>A = Zk / Xk[None, :]</code>	technical coefficients
(4)	<code>L = np.linalg.inv(I - A)</code>	Leontief inverse
(5)	<code>v, m, t = VAk/Xk, IMk/Xk, TAXk/Xk</code>	coefficient vectors
(7)–(8)	<code>E = v[:,None]*L; imp = m@L; tax = t@L</code>	embodied content
(12) (rows)	<code>S; R = vstack([S@E, imp, tax])</code>	source aggregation
(11)–(12) (cols)	<code>w = Xk[cset]/Xk[cset].sum(); R[:,cset]@w</code>	output-weighted columns
(13)	<code>M = 100*F / F.sum(axis=0)</code>	percentages
(14)–(15)	<code>build_difference()</code> (nanmedian; <code>pak - bench</code>)	benchmark & Δ

A The exact 35 $\rightarrow$ 15 sector mapping

This is the mapping implemented in `va_decomposition_workflow.py` (the MAP dictionary and the PAPER list). It refines the 11-sector paper aggregation documented in `IOT_sector_mapping.docx` in exactly one place: the single “Non-food manufacturing” bucket is split into five using sectors (Textiles; Petrol. & Chem.; Machinery & Elec.; Transport eq.; Other mfg) so the heatmaps show manufacturing detail. Every non-manufacturing sector is identical to that document; “Other mfg” is the residual and also absorbs mining.

#	ADB IOT sector (35-sector table)	Paper sector
1	Agriculture, hunting, forestry, and fishing	Agriculture
2	Mining and quarrying	Other mfg

#	ADB IOT sector (35-sector table)	Paper sector
3	Food, beverages, and tobacco	Food
4	Textiles and textile products	Textiles
5	Leather, leather products, and footwear	Other mfg
6	Wood and products of wood and cork	Other mfg
7	Pulp, paper, paper products, printing, and publishing	Other mfg
8	Coke, refined petroleum, and nuclear fuel	Petrol. & Chem.
9	Chemicals and chemical products	Petrol. & Chem.
10	Rubber and plastics	Petrol. & Chem.
11	Other nonmetallic minerals	Other mfg
12	Basic metals and fabricated metal	Other mfg
13	Machinery, nec	Machinery & Elec.
14	Electrical and optical equipment	Machinery & Elec.
15	Transport equipment	Transport eq.
16	Manufacturing, nec; recycling	Other mfg
17	Electricity, gas, and water supply	Government
18	Construction	Construction
19	Sale, maintenance, and repair of motor vehicles; retail sale of fuel	Trade
20	Wholesale trade and commission trade	Trade
21	Retail trade; repair of household goods	Trade
22	Hotels and restaurants	Other services
23	Inland transport	Transport
24	Water transport	Transport
25	Air transport	Transport
26	Other supporting and auxiliary transport activities	Transport
27	Post and telecommunications	Prof. services
28	Financial intermediation	Prof. services
29	Real estate activities	Construction
30	Renting of machinery and equipment and other business activities	Prof. services
31	Public administration and defense; compulsory social security	Government
32	Education	Education
33	Health and social work	Health
34	Other community, social, and personal services	Other services
35	Private households with employed persons	Other services

The decomposition uses these $G = 15$ paper sectors as the *using* sectors (columns of $\mathbf{M}$) and as the domestic *source* rows, plus two source-only rows – *Imported inputs* and *Net taxes* – giving the 17×15 matrix of Section 4.

References

- [1] Leontief, W. (1936). Quantitative input and output relations in the economic system of the United States. *Review of Economics and Statistics*, 18(3), 105–125. [*Origin of the input–output model and the inverse used here.*]
- [2] Miller, R. E., & Blair, P. D. (2009). *Input–Output Analysis: Foundations and Extensions* (2nd ed.). Cambridge University Press. [*The standard textbook; Chs. 2 and 6 cover $\mathbf{A} = \mathbf{Z}\hat{\mathbf{x}}^{-1}$, the Leontief inverse, and value-added multipliers.*]
- [3] Hummels, D., Ishii, J., & Yi, K.-M. (2001). The nature and growth of vertical specialization in world trade. *Journal of International Economics*, 54(1), 75–96.
- [4] Johnson, R. C., & Noguera, G. (2012). Accounting for intermediates: Production sharing and trade in value added. *Journal of International Economics*, 86(2), 224–236. [*Value-added exports via $\hat{\mathbf{v}}(\mathbf{I} - \mathbf{A})^{-1}$.*]
- [5] Koopman, R., Wang, Z., & Wei, S.-J. (2014). Tracing value-added and double counting in gross exports. *American Economic Review*, 104(2), 459–494. [*Canonical decomposition of gross exports into domestic and foreign value-added terms.*]
- [6] Wang, Z., Wei, S.-J., & Zhu, K. (2013). Quantifying international production sharing at the bilateral and sector levels. *NBER Working Paper* No. 19677.
- [7] Los, B., Timmer, M. P., & de Vries, G. J. (2015). How global are global value chains? A new approach to measure international fragmentation. *Journal of Regional Science*, 55(1), 66–92. [*Hypothetical-extraction on the same $\mathbf{A}$.*]
- [8] Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., & de Vries, G. J. (2015). An illustrated user guide to the World Input–Output Database: the case of global automotive production. *Review of International Economics*, 23(3), 575–605. [*Accessible walkthrough of the value-added multiplier used above.*]
- [9] Antràs, P., & Chor, D. (2018). On the measurement of upstreamness and downstreamness in global value chains. In L. Y. Ing & M. Yu (Eds.), *World Trade Evolution: Growth, Productivity and Employment* (pp. 126–194). Routledge.

These references are provided as reliable entry points to the literature; please verify exact titles, years and page numbers against the originals before formal citation.